

本周的主要工作是实现时间序列查找结果的可视化。

1、聚类

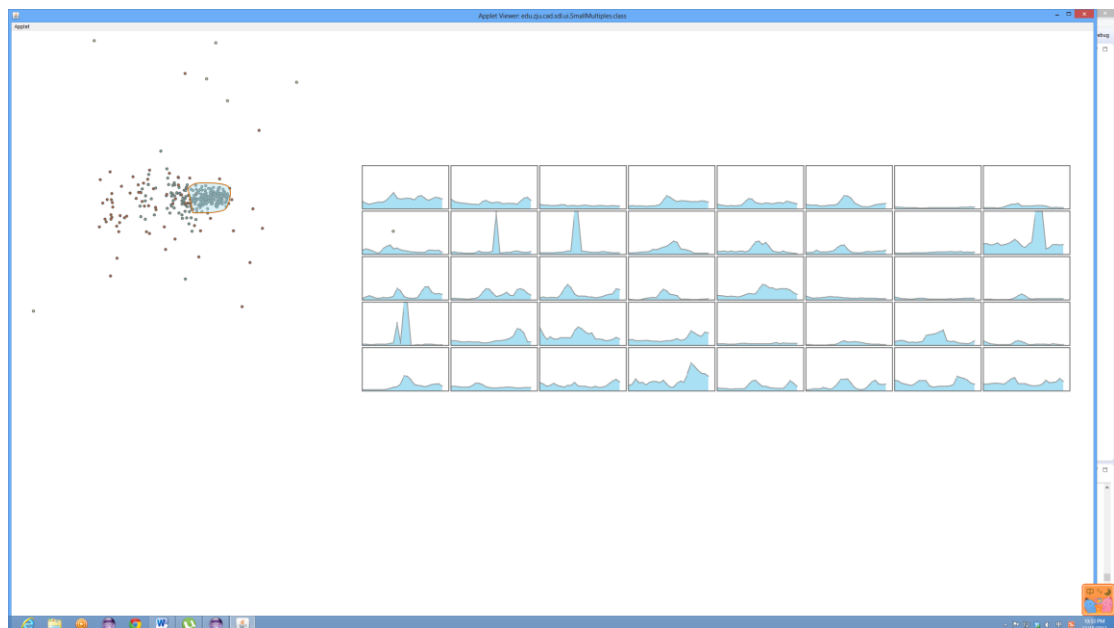
由于搜索的结果较多，尝试对数据进行聚类，聚类采用了简单的 KMeans 算法，欧氏距离作为度量，从结果上看大多数聚类在第二个分类中，其他三个分类数据较少(k 取 4)。

2、投影

投影则采用了 MDS 算法，由于 MDS 是全局投影，投影点的位置并不固定，导致每次可视化的结果略有不同。

3、套索交互工具

从投影结果看，聚类结果有一些明显的模式，使用套索工具选择聚类紧密的点，可以查看这些数据是否相似，选择后的视图如下图。从图中可以看出，聚类、投影之后的效果并不是很理想，数据的形态有很多不一致。



产生这种结果的原因可能有以下几点：

- 1) 欧氏距离并不能反映形态，需要形态一致还需要新的度量方法
- 2) 由于对形态的值设定范围较大，如 Up 为斜率 0.2—1.0，所以尖峰也会被当做 Up，需要定义更多的形态

4、可视化方法

从搜索任务上来看，我们的目标是通过相似时间序列来找到目标污染物以及污染物的相关性，相似时间序列只是中间过程，最终需要通过相似时间序列的 rank 来确定结果。如何确定 rank，目前还没有确切的方法。或者是定义一个 rank 计算规则，或者是由用户通过交互自己确定 rank 排列，这样的话可以提供一个 calendar view，可以参考下图。特别是关于时间部分的交互，只是在每个格子中可以用时间序列代替而不是数字，可以让用户更清楚地看到结果，自己对结果做评判。

